PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/04, C12N 15/63
C12Q 1/68, G01N 27/26

A1

(11) International Publication Number: WO 94/07907

(43) International Publication Date: 14 April 1994 (14.04.94)

(21) International Application Number: PCT/US92/08284

(22) International Filing Date: 25 September 1992 (25.09.92)

(71) Applicant: ZOOGEN, INCORPORATED [US/US]; 1105 Kennedy Place, Suite 4, Davis, CA 95616 (US).

(72) Inventors: HALVERSON, Joy, Lynn ; DVORAK, Jan ; 2442 Elendil Lane, Davis, CA 95616 (US).

(74) Agents: ROWLAND, Bertram, I. et al.; Flehr, Hohbach, Test, Albritton & Herbert, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111-4187 (US).

(81) Designated States: AU, CA, FI, JP, NO, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE).

Published
With international search report.

(54) Title: AVIAN SEX IDENTIFICATION PROBES

(57) Abstract

Avian cDNA nucleotide sequences for use as hybridization probes are provided. Probes derived from the sequences find use in defining the sex of a bird, usually by producing hybridization patterns that are sex-specific for most avian species. For a given avian species the probes may hybridize to both Z and W chromosomes so as to differentiate between the two chromosomes on the basis of restriction fragment length polymorphisms. Alternatively, the probes may hybridize exclusively to one of the two sex chromosomes in some species. The hybridization probes taught are also useful for the isolation of other nucleotide sequences that may be used to generate sex-specific hybridization patterns.

AVIAN SEX IDENTIFICATION PROBES

INTRODUCTION

Technical Field

This invention relates to the field of restriction fragment length polymorphism markers suitable for sex identification in avian species, and methods for using such markers.

Background

RFLP (restriction fragment length polymorphism) analysis has proven highly successful in mapping and discovering new genes in eukaryotic organisms. RFLP marker probes consist of polynucleotide sequences that specifically hybridize to a region of the chromosome. These chromosomal regions of hybridization are revealed to be polymorphic between individuals of the same species when the chromosomal DNA is digested by restriction endonucleases and analyzed by hybridization analysis. Different RFLP alleles are distinguished from one another on the basis of the hybridization banding patterns produced after size separation. Genetic linkage analysis between RFLP markers and uncharacterized genes has proven to be a useful technique for isolating and mapping uncharacterized genes of interest. Although the RFLP marker has most frequently been used to identify the chromosomal disruptions responsible for genetic diseases, it is also of interest to use RFLP markers in deciphering

other complex genetic regulatory questions, such as sexual development in animals.

While the genetics and biochemistry of sex determination in mammals has been the subject of extensive scientific investigation, similar studies have not been carried out with respect to birds. This is surprising in view of the commercial importance of numerous avian species e.g., chickens and turkeys. It is of interest to provide research tools useful for deciphering the complex process of genetic sex determination in birds. Sex-specific genetic markers are of particular interest. Such markers may be used to sexually identify immature birds prior to the development of gender specific morphological differences. Early sexual identification is an important consideration when breeding those birds that become sexually mature prior to the development of external sexual characteristics. Accordingly, there is interest in providing methods for preventing undesired matings by permitting gender identification (and gender separation) prior to the development of sexual maturity.

Genetic sexual identification is also useful in the breeding of rare bird species with unidentified secondary sexual characteristics; captive breeding programs may thus be effectively organized.

Sex chromosomes, as opposed to autosomal chromosomes, differ with respect to one another in size and genetic composition. Thus, some regions of one sex chromosome contains genes which have no corresponding allele on the other sex chromosome.

One of the principal ways in which the sex chromosomes of birds differ from man and other mammals is that the

female bird is the "heterogametic" sex, having Z and W sex chromosomes. In mammals, the male is the "heterogametic" sex having both X and Y chromosomes whereas the female is "homogametic" having two X chromosomes.

It is of interest to provide RFLP genetic markers suitable for the identification of DNA regions which are diagnostic of the sex of the bird, where the DNA regions may be common to the Z and W chromosomes such that RFLP's unique to each sex chromosomes may be detected; or the DNA region is unique to one of the sex chromosomes, so that the amount or presence of the RFLP will determine the sex of the bird. Such genetic markers permit the identification of the chromosomally specified sex of an individual bird based on analysis of a DNA preparation derived from the bird.

Relevant Literature

Halverson J., Dvorak, J., Flammer, K. 1985. A new method of avian sex determination- identification of the W body by C-banding of erythrocytes. Proceedings of the Annual Meeting of the Association of Avian Veterinarians. 1985. Boulder, Colorado.

Halverson, J., Rauen, K., 1988. The molecular approach to poultry breeding. Proceedings of the Thirty-seventh Western Poultry Disease Conference and Molecular Biology Workshop. 1988. Davis, California.

SUMMARY OF THE INVENTION

Avian nucleotide sequences for use as a source of hybridization probes are provided. Probes derived from these sequences, Tsex (also designated as pMg1) and related sequences, are used in hybridization for sex identification of many avian species. Procedures are given for using the probes so as to sexually identify

individual birds. For a given avian species, the probes may hybridize to both Z and W chromosomes allowing for differentiation between the two chromosomes on the basis of restriction fragment length polymorphisms. Alternatively, the probes may hybridize exclusively to one of the two sex chromosomes in some species, thus permitting gender identification on the basis of sex-specific hybridization intensity. The probes provide a means for identifying the gender of a bird without reliance on morphological characteristics. The probes also find use in recovering and identifying nucleotide sequences homologous to Tsex and sex-specific nucleotide sequences that are genetically linked to genomic Tsex or Tsex-homologous sequences.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Novel nucleic acid sequences and methods for using the sequences are provided for determining the chromosomally specified sex of individual birds. The method is based on the ability of hybridization probes derived from the nucleic acid sequences to hybridize with sequences from the sex chromosomes of birds being analyzed for sexual identification. The nucleic acid sequences provided for, namely Tsex, derivatives thereof, and equivalent sequences, are homologous to nucleic acid sequences present on both or one of the Z and W chromosomes of most birds tested. However, there do exist avian species in which Tsex-homologous sequences are only present on one of the two sex chromosomes. For bird DNA, the term homologous when applied to nucleic acid sequences intends nucleic acid sequences capable of hybridizing to each other at a stringency of at least about 25°C below the T_m (T_m is the temperature at which about half the nucleic acid strands are denatured). In referring to "Tsex homologous sequences" it is intended to include the Tsex sequence itself. Most generally used

hybridization protocols may be used with Tsex-derived probes. See Molecular Cloning: A Laboratory Manual, second edition (1989), Sambrook et al., Cold Spring Harbor Laboratories Press, Cold Spring Harbor, N.Y., for examples of such protocols.

The subject invention permits the identification of the chromosomally specified sex of a given bird based on a hybridization analysis of genomic DNA extracted from the individual bird being gender tested. Suitable birds for chromosomal identification by Tsex hybridization have Tsex-homologous sequences present on one or both sex chromosomes.

The subject invention provides a means for identifying the gender of avian species without relying on morphological sexual characteristics. Thus the subject invention may be used to identify the sex of birds when secondary (external) sexual characteristics are absent or unknown. The identification of sex in the absence of secondary sexual characteristics is useful because in commercial breeding operations it is advantageous to separate sexually mature birds from one another in order to prevent undesired matings. This separation may be difficult to achieve because some avian species become sexually mature prior to the development of obvious external sexual differences. Thus, by providing for a simple and reliable means of gender identification prior to the development of external sexual characteristics in birds, the subject invention finds use in commercial bird breeding. The subject invention finds use in determining the gender of rare bird species in which the gender specific external characteristics are unknown. By permitting the creation of breeding pairs, the subject invention provides a rational means of breeding rare birds.

Hybridization probes derived from Tsex find a variety of uses in addition to their use in sexual identification. Tsex-derived hybridization probes may be used to isolate Tsex-homologous sequences. Such isolation may be achieved by screening recombinant DNA libraries prepared from avian DNA (or cDNA) from species other than turkey (the Tsex sequence being derived from a turkey cDNA). In addition, these Tsex probes may be used to isolate full length cDNAs for Tsex and the Tsex genomic sequences. Similarly, Tsex-derived hybridization probes may also be used in chromosome walking or jumping techniques to isolate coding and noncoding sequences chromosomally proximal, though not necessarily adjoining, to Tsex-homologous sequences. Tsex sequence derived probes may be employed with chromosome walking and jumping techniques as described in a number of commonly available publications, e.g., Sambrook et al., supra.

Additional uses of the subject invention are found in the isolation of sex determination mutations in avian species. Mutations affecting sex determination have found use in manipulating progeny phenotype in the controlled breeding of numerous animals (e.g., linked X chromosomes, autosomal-sex chromosome translocations). The subject invention facilitates the discovery of sex determination mutants by allowing investigators to determine the sexually specified genotype of mutant birds with questionable sexual morphology.

The Tsex sequence was obtained from a cDNA library prepared from turkey embryonic poly(A)mRNA. The Tsex sequence is 959 base pairs in length. The sequence of Tsex is displayed in Table 1.

Table 1

```
1   5'AACAGCATCT GATGCTGCAC CCCTTCAGTA TCTGGCTCCC TACTCAGGCT
51    GCTCCATGGG GGAATACTTC AGAGACAATG GGAAACATGC ATTGATCATC
101   TACGATGACT TGTCCAAACA GGCTGTTGCC TACCGTCAGA TGTCTCTGCT
151   GCTGCGTCGT CCGCCTGGCC GTGAAGCTTA CCCAGGTGAT GTGTTCTACC
201   TGCACTCTCG CCTGCTGGAG AGAGCAGCTA AAATGAATGA TTCCTTTGGA
251   GGAGGCTCTC TGACTGCTTT GCCCGTCATT GAAACTCAGG CTGGTGATGT
301   GTCTGCTTAC ATTCCAACCA ATGTCATCTC CATCACTGAT GGACAGATCT
351   TCTTGGAAAC TGAACTGTTC TACAAAGGTA TCCGTCCAGC CATCAACGTT
401   GGTCTGTCTG TGTCCCGTGT GGGTTCTGCT GCTCAGACCA GGGCAATGAA
451   ACAGGTGGCT GGTACCATGA AGCTGGAGCT GGCTCAGTAC CGTGAAGTGG
501   CTGCCTTTGC TCAGTTTGGG TCTGATTTGG ATGCTGCCAC ACAACAGCTG
551   CTGAATCGTG GTGTGCGTCT GACAGAGCTC CTGAAACAAG GACAGTATGT
601   TCCCATGGCT ATTGAGGAAC AGGTTGCAGT CATCTATCGT GGTGTAAGAG
651   GTCACTTGGA CAAGCTGGAG CCCAGCAAAA TCACTAAATT TGAGAGTGCT
701   TTCCTGGCTC ATGTACTGAG CCAGGACCAG GCCCTCCCTC TCCACCATCA
751   GGACTGAAGG GAAGATCTCT GACCAGACGG AAGCTAAGCT GAAGGAAATA
801   GTCACAAATT TCCTATCTAC TTTTGAGGCA TAAACTCATT ATCTGTTCAA
851   ACAGACCAGG CTGTTTTTGT TGTTACGTGC TTTGCCTCCA TCAAAGACCT
901   AAACGTATCG AGTGCTTGAA TGTACAGATC TCACTGAGAA TAAAAGTTTC
      CATGTAAAA3'
```

Sequence displayed from position 1 to end (position 959)

The Tsex sequence may be used for the production of a variety of nucleic acid hybridization probes, also referred to as Tsex-derived probes.

Hybridization probes based on the Tsex sequence find use for many purposes, including sex identification in birds and the isolation of Tsex-homologous sequences from a genetic library. Probes may be either single or double stranded, either RNA or DNA. Probes may be produced by <u>in vitro</u> or <u>in vivo</u> synthesis. Probes may also be produced by a combination of <u>in vitro</u> and <u>in vivo</u> synthesis. Methods of <u>in vitro</u> probe synthesis include organic chemical synthesis processes or enzymatically mediated synthesis, e.g., by means of SP6 RNA polymerase and a plasmid containing the Tsex sequence under the transcriptional control of an SP6 specific promoter. Usually, probes will have a specific complementary sequence of at least 12 nucleotides, more usually at least 14 nucleotides and preferably greater than 50 nucleotides and, more preferably having the entire Tsex sequence.

Probes may bear either complete or partial homology to Tsex. Probes containing partial homology to Tsex will usually have less than 20% mismatch with Tsex, and preferably less than 10% mismatch with Tsex.

When performing sex determination analysis using Tsex sequence derived probes not actually containing the complete 959 base pair Tsex sequence, the probes are preferably hybridized against DNA preparations from birds of known sex in order to verify that the probes produce the desired sex-specific hybridization pattern. Similarly, when sexing previously untested species of birds with Tsex sequence derived probes, the sex-specific nature of the hybridization patterns produced with Tsex sequence derived probes are preferably verified by hybridization against birds of known sex.

Probes may be modified by conjugation to a variety of labels which allow for detection of duplex formation between the probe and its complementary target. Labels include radioactive isotopes, ligands, e.g., biotin, enzymes, fluorescers and the like. A wide variety of protocols for labeling probes and detecting duplexes formed between probes and their target hybridization sequences have been described in the literature. See for example, Berger and Kimmel, editors, Guide to Molecular Cloning Techniques (1987) Academic Press Inc., San Diego, CA.

Probe sequences may be joined to a variety of other nucleic acid sequences. Among these other nucleic acid sequences are vectors such as plasmids, cosmids, phages, and the like. By joining the probe sequence to a vector sequence, probes may be conveniently created, expanded, stored, and modified.

Nucleic acid preparations suitable for hybridization analysis with Tsex-derived probes may be isolated from any portion of a bird's body containing substantially intact nucleic acids. Preferably, nucleic acid preparation sources may be obtained without killing or injuring the bird and may be easily removed from the test bird. Preferred sources of starting material will be feather pulp, blood, or the like.

Tsex-derived hybridization probes may be used to determine the sex of an individual bird based on nucleic acid hybridization analysis. As previously discussed, hybridization analysis demonstrates that Tsex-homologous sequences are present on at least one, and usually both sex chromosomes of numerous avian species. For those species that have Tsex-homologous sequences present on both Z and W sex chromosomes, hereinafter called Tsex-double species, Tsex sequence derived hybridization probes may be used to distinguish male from female DNA preparations on the basis of the presence of at least one additional hybridization band present in the female chromosomal preparations. For those bird species having Tsex-homologous sequences present on only one sex chromosome, hereinafter called Tsex-single species, the genetically specified sex of the bird from which the DNA was prepared may be determined by employing a Tsex-specific probe and semi-quantitatively comparing the degree of hybridization of the probe to a sample with the degree of hybridization of the same probe to a standard of known sex.

Tsex-double species include, but are not limited to: turkeys, chickens, Canada geese, pheasants, zebra finches, lapwings, sandgrouse, murres, caiques (including: white-bellied and black-headed), macaws (including: blue & gold, scarlet, hyacinth, yellow-collar, hahns, greenwing, and military), red-bellied

parrots, amazons (including: Mexican red-headed, orangewing, red lored, mitred, cherry-head, yellow-nape, and blue-front), cockatoos (including: rose-breasted, goffins, umbrella, mollucan, and sulfur-crest), lorries, cockatiels, budgerigars, and rosellas.

Tsex-single bird species include but are not limited to: African grey parrots, Eclectus parrots, Lovebird, and Conures (including: Halfmoon, Patagonian, and Nanday).

Some bird species are not amenable to sex identification with Tsex-derived probes with the restriction endonucleases employed, but may be capable of identification with an appropriate endonuclease. Tsex-derived probes hybridize to sequences on both the W and the Z chromosomes in these species; however, the hybridization patterns produced do not vary between the sexes. The only species tested that are not currently amenable to sex identification with Tsex-derived probes are bald eagles, red tail hawks, ostriches, penguins, cinerous vultures, and Russell Griffon vultures.

The preferred method of sex identification by hybridization analysis with Tsex-derived nucleic acid hybridization probes requires the immobilization of target hybridization sequences onto a solid support. Avian sex identification employing the process of nucleic acid hybridization with Tsex-specific probes may utilize solid immobilization supports suitable for hybridization analysis with most generally recognized nucleic acid hybridization procedures. Suitable support membranes capable of binding nucleic acids include nitrocellulose, Nytran™, Zetaprobe™ and the like. A variety of protocols for immobilizing nucleic acids to membranes have been described in the literature. See for example, Berger and Kimmel, supra.

The immobilization procedure may be mediated by capillary transfer, electrophoretic transfer, vacuum transfer and the like. The amount of nucleic acid transferred to the solid support membrane will generally be in the range of about .01 μg to 20 μg, preferably in the range of about .05-5 μg, per sample for analysis.

Chromosomal DNA preparations suitable for hybridization with Tsex-derived probes may be digested with restriction endonucleases prior to the transfer of the nucleic acids to an immobilizing support, usually after size separation where the sex is determined based on the difference in size of one or more fragments. Prior to immobilization, chromosomal DNA preparations from Tsex-double species must be digested with at least one restriction endonuclease, whereas chromosomal DNA from Tsex-single species are preferably, though not essentially, digested with at least one restriction endonuclease. Useful restriction endonuclease are selected for their ability to produce a sex-specific hybridization pattern when hybridized against Tsex-derived probes. Where restriction sites are present in the Tsex- or homologous DNA, a number of fragments which bind to the probe may be produced. Digestion with suitable restriction endonucleases produces chromosomal fragments with sizes capable of being separated by electrophoresis.

When DNA preparations for gender determination analysis are subjected to restriction endonuclease digestion, the DNA preparations are size separated by electrophoresis, and transferred <u>in situ</u> from the gel electrophoresis separation medium to a nucleic acid binding support. Electrophoretic separation should proceed to a point where the degree of chromosome digest fragment separation achieved by electrophoresis

is sufficient to separate Tsex-homologous chromosomal fragments from one another to an extent capable of being detected by hybridization analysis. Suitable gel electrophoresis separation media include agarose, polyacrylamide, mixtures thereof, and the like, at a concentration suitable for the separation of nucleic acid fragments to an extent sufficient to reveal size differences between Tsex-homologous DNA fragments. The in situ transfer of nucleic acids to a binding membrane may be achieved by any of the standard procedures for hybridization analysis. Exemplary, but not exclusive of such procedures are Southern blotting and electroblotting.

After immobilization to a solid support, the samples are hybridized against Tsex-derived hybridization probes. Duplexes formed between the labeled probe and the Tsex-homologous sequences in the samples for analysis and any hybridization standards present are subsequently visualized by methods appropriate to the probe label.

When performing chromosomal sex analysis on Tsex-double species, male and female chromosomal DNA hybridization sex standards are optionally, although preferably, present. By hybridization standards, it is intended chromosomal DNA from an individual of known sex and of the same species as the individual undergoing sex determination analysis; the standard is also present in a quantity essentially the same as the DNA sample for analysis.

Male and female DNA samples from Tsex-double species may be distinguished from one another by the presence of at least one hybridization band found in the chromosomal preparation of one sex that is not found in preparations of the other sex. Since, frequently one

obtains a plurality of bands of different sizes which bind to the probe, all that is required to distinguish the sex of the subject bird is that there be a band of a size characteristic of a particular sex, where the absence of the band is characteristic of the other sex. Furthermore, the intensity of the hybridization bands in common between male and female chromosomal DNA preparations have greater (usually approximately double) hybridization intensity in male derived samples because of the double dose of target DNA from the two Z chromosomes. Thus, by comparing DNA samples for analysis with each other, or comparing the samples with known male and female standards of the same species, the sex of the individual animal providing the DNA for analysis may be determined.

Sex identification in Tsex-single species is performed essentially the same as in Tsex-double species; however, the restriction endonuclease digestion and electrophoresis separation steps are optional (although preferably performed) rather than required as with the analysis of Tsex-double species. When analyzing Tsex-single species, all DNA samples for analysis and hybridization standards are necessarily present in approximately the same quantity so that the intensity of hybridization produced may be accurately compared among samples and hybridization standards.

In Tsex-single species where Tsex-homologous sequences are only present on the Z chromosome, probe hybridization with chromosomal preparations from males displays greater (approximately double) hybridization intensity than probe hybridization with equal quantities of DNA from female members of the same species.

Tsex-derived probes also find use in isolating other sex-specific sequences. As previously discussed, Tsex-derived probes are useful for sex identification because hybridization with these probes reveals sex-specific hybridization patterns. Since the Tsex probes which hybridize to the W and Z chromosomes are the same sequence, it is apparent that the sex-specific nature of the hybridization must be attributable to chromosomal sequences other than Tsex itself (or Tsex-homologous sequences in species other than turkey). Thus on the chromosome, Tsex-homologous sequences lie close to nucleic acid sequences that differ from one another in a sex-specific manner. These as yet unidentified sex-specific sequences may be detected by employing conventional library screening techniques, including chromosome walking (Bender et al., J. Mol. Biol. (1983) 168:17) and chromosome jumping (Poustka et al. Nature (1987) 325:323), to recover sequences proximal to Tsex-homologous sequences, in which at least the first round of library screening employs a Tsex sequence derived probe. These library screening techniques are discussed at length in Berger and Kimmel, supra and Sambrook et al., supra.

Genetic libraries for screening with Tsex-derived probes may be prepared in plasmid, cosmid, or YAC (yeast artificial chromosome) vectors or the like. Portions of nucleotide sequences isolated in a first round of screening that are not homologous to Tsex may be labeled and used as probes for a second round of library screening. The process of repeatedly screening a library with newly isolated sequences may be repeated as desired so as to "walk" or "jump" down the chromosome. Progression down the chromosome from the Tsex homologous region may proceed in either direction. The region of the chromosome analyzed by "walking" techniques and the like may extend for a distance

greater than 5000 kb to either side of the Tsex region (or Tsex-homologous region), however preferred distances for walking are in the range of about 100-1000 kb. Chromosomal regions of interest also include introns within the Tsex-homologous genomic sequence. Nucleotide sequences not homologous to Tsex that are isolated during "walking" may be screened for their ability to produce sex-specific hybridization patterns when used as hybridization probes. It is also of interest to engage in chromosome walking for the purpose of isolating sex-specific nucleotide sequences in those bird species that have non-polymorphic Tsex-homologous sequences on both the W and the Z chromosomes, e.g., penguins.

For an example of using Tsex to isolate additional sex-specific nucleotide sequences, consider the case of screening a turkey genomic library with a labeled Tsex DNA sequence as a probe. Plasmids from individual clones recovered in the screening may then be restriction mapped and regions of the recovered plasmids not homologous to Tsex isolated. The isolated regions may then be labeled for use as probes in a second round of library screening. Sequences recovered from library screening may also be tested for use as sex-specific probes by labeling the newly isolated sequences and hybridizing the labeled sequences against Southern blots containing restriction digested genomic DNA from both male and female birds (of the same species). Sequences that reveal sex-specific hybridization pattern polymorphisms may find use in sex identification protocols similar to those employing Tsex-derived probes.

Besides using the nucleic acid as an identifier of sex, the expression product may also be used to identify sex. To the extent that the expression products of the

two sequences differ as to epitopes, antisera or monoclonal antibodies may be prepared which will distinguish between the two alleles, when the Tsex or related gene is present on both chromosomes. Where the gene is only on one chromosome, the amount or presence of the protein will indicate the sex of the bird. Various immunoassays may be used to detect the proteins, using radioisotopes, enzymes, chromophores, fluorophores, chemiluminescers, or the like, as detectable labels.

Kits may be provided with probes and standards for determining the sex of birds. Thus, by carrying out the assay with the sample and having genomic DNA which may be processed in the same manner, the results may be compared directly. Thus, the kit would comprise, one or more probes, usually one probe, generally labeled, and genomic DNA for one or both sexes of the bird of interest for processing in the assay and comparison with the sample DNA.

The invention now being generally described, the same will be better understood by reference to the following examples which are provided for purposes of illustration only and are not to be considered limiting of the invention unless so specified.

EXAMPLES

RNA Isolation

A cDNA library was prepared from poly A mRNA isolated from turkey embryos. All RNA isolation procedures and manipulations were performed in RNAse free laboratory articles. In order to remove RNAse contamination, glassware was baked at 180° to 190°C overnight and plasticware was treated with a 0.2% diethylpyrocarbonate (DEPC) solution.

8.3 gms of 3, 4, and 5 day old turkey embryos were pooled together and placed in a lysis buffer (7 ml of buffer/1 gm tissue). The lysis buffer is composed of:

4M Guanidinium isothiocyanate
.2M Tris pH 7.5
.01M EDTA
5% β mercaptoethanol (w/v)

The tissue was homogenized in a Polytron™ homogenizer at setting "5" for 1 minute. The homogenate was strained through a cheese cloth and a myra cloth filter (a double layer in which the homogenate first passes through the cheese cloth). The filtrate was then centrifuged at 10,000 x g at 25°C for 10 minutes. The supernatant was removed and N lauryl sarcosine was added until a final concentration of 0.5% was reached. Seven volumes of 4 M LiCl were then added. The supernatant was then allowed to sit for 15-20 hours at 4°C. The solution was then centrifuged at 10,000 x g for 90 minutes at 4°C. The supernatant was removed, and the pellet was resuspended in 100 ml of 3 M LiCl. The solution containing the resuspended pellet was then centrifuged at 10,000 x g for 60 minutes at 4°C and the supernatant removed. The pellet was then dissolved in a solubilization buffer by vortexing. RNA solubilization buffer: 0.1% SDS, 0.01 M pH 7.5 Tris, EDTA pH 8 0.5 mM (about 50 ml buffer/10 gm sample). The RNA was then extracted with an equal volume of phenol : chloroform : isoamyl alcohol (25:25:1) solution. The aqueous phase of the extraction was saved and a 3 M potassium acetate solution was added to make a solution with a final concentration of 0.2 molar potassium acetate. 2.2 volumes of 95% ethanol were then added. The solution was dispensed into 30 ml Corex™ tubes. The Corex™ tubes were stored at -20°C overnight. The tubes were then centrifuged at 10,000 x g for 30 minutes at 4°C. The supernatant was removed and the pellet was washed with a 70% ethanol solution.

The tube was allowed to drain and then subsequently dried in a vacuum oven for 20-60 minutes. The pellet was then resuspended in sterile deionized water.

Isolation of Poly A mRNA

Poly A mRNA isolated from the total embryonic turkey RNA preparation by oligo(dT)-cellulose was equilibrated with about 2 ml of loading buffer (loading buffer; 20 mM Tris pH 7.5, 0.1 mM EDTA, 0.5 M LiCl, 0.1% SDS). The oligo (dT)-cellulose slurry was poured into a 1.0 ml siliconized Pasteur pipette (plugged with siliconized glasswool). The column was washed with 3 volumes each of (1) sterile deionized water, (2) 0.1M NaOH, 5 mM EDTA, (3) sterile deionized water. The column was washed with water until the column effluent had a pH of less than 8.0. The column was then washed with 5 volumes of loading buffer. An aqueous solution containing the total RNA isolated from turkey embryos was heated to 65°C for 5 minutes. Equal volumes of 2x loading buffer was added to the RNA solution and the solution was allowed to cool to room temperature. The solution was then applied to the top of the oligo(dT)cellulose column. The column effluent was collected and heated again to 65° and reapplied to the top of the column. The column was then washed with 5 mls of loading buffer. 1 ml aliquots of column effluent were collected. 5 volumes of 0.1 M LiCl buffer (20 mM Tris pH 7.5, 1 mM EDTA, 0.1 M LiCl, 0.1% SDS) were then added to the column. The mRNA fraction was then eluted with 5 volumes of eluting buffer (1mM Tris pH 7.5 1 mM EDTA, 0.05% SDS). 1.0 ml aliquots of column effluent were collected. The column was then rinsed with the deionized water, until the concentration of RNA in the effluent was approximately 0. The A_{260} of each column fraction was determined in order to measure RNA concentration. The fractions were

then stored at -70°C in 70% ethanol and 0.2 M potassium acetate.

cDNA Cloning

The embryonic turkey cDNA library was generated using the methods and vectors described by Alexander et al. "Dimer - Primer cDNA Cloning", Methods in Enzymology 154, Academic Press (1987). The cDNA was cloned into the SstI, site of pARC 7. Plasmid pARC 7 was digested with SstI and poly (dT) tailed using terminal deoxynucleotidyl transferase. Isolated embryonic turkey mRNA was allowed to anneal to the poly-dT tails via the poly-A tails of the mRNAs. After annealing to the poly-dT tail primers, the first strand of cDNA synthesis was catalyzed using murine Moloney leukemia virus reverse transcriptase. The first DNA strand was then poly-dG tailed using terminal deoxynucleotidyl transferase. The modified vector bearing cDNA strands on both termini was then digested with the restriction enzyme BamHI. A BamHI linker with a poly-dC tail was then annealed to the poly-dG tail at 42°C. The mixture was then cooled, permitting the constructs to circularize through the annealing of the BamHI sticky ends. Circularization was completed by the use of T4 DNA ligase. The RNA strand still remaining in the construction was removed by mixing the RNA-DNA duplex with RNase H and replacing the RNA with DNA using DNA polymerase I. The remaining nick was then closed with T4 DNA ligase and the constructs were transformed into E. coli strain DH5α.

Isolation of pTsex

Clones from the turkey embryonic cDNA library were selected at random and the plasmids within them prepared by a plasmid miniprep procedure (Birnboim and Doly, Nucleic Acids Research (1979) 7:1513). The turkey embryonic cDNA library plasmids were then

screened by digestion with either PstI or XbaI, resulting in the release of the insert portion of the plasmid, and subsequent separation of the vector from the insert by agarose gel electrophoresis. Clones bearing inserts of greater than about 400 base pairs were subjected to further screening.

Further screening was carried out by radioactively labeling the cDNA portion of the isolated clones and using the resulting radioactive probe to screen Southern blots containing restriction digested chromosomal adult turkey DNA from both male and female turkeys. Prior to radioactive labeling, the inserts from the clones for further analysis were subjected to digestion with either PstI or XbaI, followed by separation of the digestion fragments from one another by agarose gel electrophoresis. The plasmid insert was isolated from the gel by electroelution. The insert was then radioactively labeled by nick translation. Nick translation was performed by placing the following mixture in a 1.5 ml eppendorf tube: 0.5 μg of insert DNA, 2.5 μl of 10x nick translation buffer (500 mM Tris pH 7.8, 50 mM MgCl, 100 mM β mercaptoethanol, 100 μg/ml BSA), 2.5 μl 2mM dGTP, 2.5 μl 2 mM dTTP, 2.5 mM 2.0 mM dCTP, 5 μl dATP 3000 curies/mM ^{32}P, 2 μl DNAse I (1U/100μl), 1 U DNA polymerase I, H_2O to 25 μl total volume. The mixture was incubated 75 minutes at 15°C. 8 μl .25 M EDTA was added, and the mixture was incubated 65°C for 10 minutes to stop the reaction. 1 μl of 10 mg/ml sonicated salmon sperm DNA was then added in order to increase recovery of the probe. The unincorporated nucleotides were removed by gel filtration through a G-75 Sephadex column.

The ^{32}P labeled inserts were then used as probes to hybridize with a Southern blot containing restriction digested chromosomal DNA prepared from turkey blood.

Southern blots used for the screening of the embryonic turkey cDNA library contained chromosomal DNA from eight turkeys, four males and four females. 5 μg specimens of chromosomal DNA isolated from the blood of three male and three female turkeys were separately digested with BamHI and EcoR I. Also a single pair of chromosomal DNA preparations from male and female turkeys was digested with HindIII. The chromosomal digest fragments were separated by agarose gel electrophoresis and transferred to a Zetaprobe (BioRad) filter membrane by Southern blotting. All potential gender identification markers were tested against Southern blots containing this combination of restriction digested turkey chromosomal DNA.

The pre-hybridization solution (see section on hybridization conditions for the composition of this solution) was added to a polyethylene baggy containing the Southern blot and allowed to incubate at 65°C overnight. The pre-hybridization solution was removed, and hybridization solution added. The hybridization solution was essentially the same as the pre-hybridization solution with the exception that the hybridization solution contained all of the probe labeled in the nick translation reaction (denatured at 95°C for 10 minutes). The hybridization was allowed to proceed for 36 hours at 65°C.

After hybridization, the membranes were washed by the following procedure. The first wash was in a solution consisting of 1xSSC, 1% SDS, at 45°C for 10 - 15 minutes. The blot was then washed in a solution of 0.1xSCC, 1% SDS, 0.1% sodium pyrophosphate, at 65°C for 40 minutes. After washing, the blot was blotted dry on Whatman™ 3 MM paper, wrapped in saran wrap, and exposed to Kodak™ XAR X-ray film for three days. The X-ray film was subsequently analyzed for differences in

the banding pattern between male and female chromosomal DNA preparations digested with the same restriction enzyme. The desired sex-specific probes should produce no variation in the banding patterns between DNA samples from members of the same sex.

The third embryonic turkey cDNA library clone tested by the above described procedure revealed a sex-specific hybridization pattern. Hybridization with this probe, named the Tsex sequence, with all three male turkey chromosomal DNA preparations digested with BamHI has revealed identical hybridization bands of 19.5 kb and 11.5 kb, whereas hybridization of pTsex with all three BamHI digested chromosomal DNA preparations from female turkeys resulted in the formation of hybridization bands with sizes of 19.5 kb, 13.5 kb, 11.5 kb, 7.3 kb, 6.4 kb and 5.5 kb.

EcoRI digested male chromosomal DNA probed with Tsex resulted in the formation of hybridization bands with sizes of about 13.5 kb, 3.9 kb and 3.4 kb whereas hybridization with EcoRI digested female turkey DNA resulted in the formation of hybridization bands with weights of about 13.5 kb, 9.1 kb, 3.9 kb, 3.4 kb and 2.3 kb. There was no variation in the hybridization patterns produced among the female DNA samples as well as among the male DNA samples.

Male chromosomal DNA digested with HindIII produced hybridization bands with weights of about 6.5 kb and 2.5 kb, whereas hybridization of chromosomal DNA resulted in the formation of hybridization bands with weights of about 6.5 kb, 5.7 kb, 5.3 kb, 4.4 kb, and 2.5 kb.

Blood Collection

Venous blood is collected by clipping a toenail sufficiently short that the blood vessel in the toenail (which usually ends in the distal quarter of the nail) is opened. After the blood has begun to ooze rapidly, a 20 µl Unopette™ pipette is held to the nail and blood is allowed to flow into the pipette quickly by capillary action. The other end of the Unopette™ fits onto the nozzle of a squirt bottle containing 70% ethanol. The ethanol is squirted through the pipette, washing the collected blood into a 2 ml polypropylene screw cap tube which is then capped. There should be no more than a 30 second delay between blood collection and mixing of the blood with 70% ethanol. The blood is now preserved and will be useful for DNA isolation during several weeks of storage at room temperature.

Blood can also be collected by venipuncture if desired. The wing vein (brachial vein) and jugular vein are preferred. The Unopette™ pipette can be used to transfer 20 µl of blood into the 2 ml tube for preservation in ethanol as described above. If blood is to be held for any time prior to preservation, it should be immediately chilled in an ice water bath and then held on ice.

Feather Pulp

The proximal shaft of primary feathers that are partially grown contain a mucinous pulp rich in DNA. Feathers are best removed by a firm, steady pulling motion which insures that the feather is removed intact, without leaving a stump. The proximal 2 inches of the feather is cut, and placed in 5 ml of a pre-chilled solution of DNA Isolation Buffer (DIB - 0.1 M NaCl, 0.1mM EDTA, 10 mM Tris pH 7.0). The sample in buffer should be kept on ice until it can be used for DNA preparation.

INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCT/GB 96/01341

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, no. 18, 15 September 1993, WASHINGTON US, pages 8324-8326, XP002016196 R.GRIFFITHS AND B.TIWARI: "The isolation of molecular genetic markers for the identification of sex" cited in the application see the sequence of primer ss2 see page 8324, right-hand column, paragraph 1	3-5, 9-12, 14-18
X	WO,A,94 07907 (ZOOGEN,INCORPORATED) 14 April 1994 see table 2	9-12, 15-18
X	DATABASE WPI Section Ch, Week 8849 Derwent Publications Ltd., London, GB; Class B04, AN 88-348691 XP002016199 & JP,A,63 258 580 (NICHIREI KK) , 26 October 1988 see abstract	9-12, 15-18
X	DATABASE WPI Section Ch, Week 8849 Derwent Publications Ltd., London, GB; Class B04, AN 88-348692 XP002016200 & JP,A,63 258 581 (NICHIREI KK) , 26 October 1988 see abstract	9-12, 15-18
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, March 1993, WASHINGTON US, pages 2414-2418, XP002016197 V.DELMAS ET AL.: "A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain" cited in the application see the whole document	19-21, 24,25
P,X	NATURE, vol. 375, 8 June 1995, LONDON GB, page 454 XP002016198 R.GRIFFITHS AND B.TIWARI: "Sex of the last wild Spix's macaw" see the whole document	3-5, 9-12, 14-18

INTERNATIONAL SEARCH REPORT

Interr onal Application No
PCT/GB 96/01341

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12Q1/68 C07K14/465 C07K14/47 A01K67/027
C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EMBL DATABASE, Accessionnumber D14316 Sequence reference GGJF11 from G. gallus 6 August 1993. XP002016259 compare nucleotides 5-2292 with nucleotides 2221-4508 in figure 5. --- -/--	1-14,19

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 October 1996	30. 10. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cupido, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

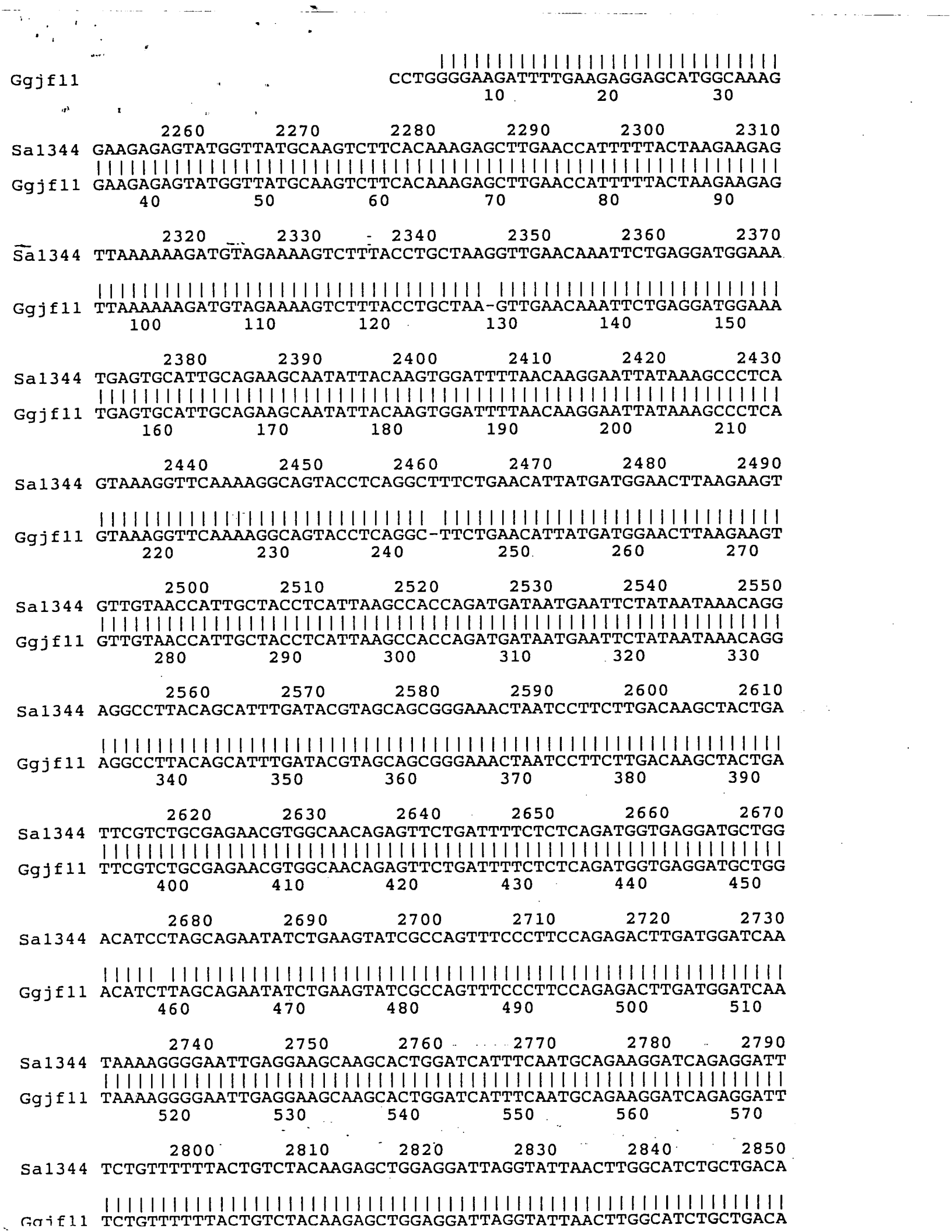
International Application No
PCT/GB 96/01341

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9407907	14-04-94	CA-A- 2124220	14-04-94
		AU-B- 662564	07-09-95
		AU-A- 2696092	26-04-94
		EP-A- 0623139	09-11-94

1/1 - (C) WPI / DERWENT
AN - 88-348691 [49]
AY - 88
XA - C88-154099
XP - N88-264155
DC - B04 C03 D16 P14
FS - CPI
PA - (NICH-) NICHIREI KK
PN - JP63258580 A 881026 DW8849
PR - JP870091976 870416
IC - A01K67/00 ; C07H21/04 ; C12N15/00 ; C12Q1/68
MC - B04-B03B B04-B04A1 B11-C08E B12-K04A6 C04-B03B C04-B04A1 C11-C08E C12-K04A6 D05-H09 D05-H12
M1 - [01] M423 M710 M781 M903 P831 Q233 V753
UP - 8849
UCC - 8849
AP - JP870091976 870416
XIC - A01K-067/00 ; C07H-021/04 ; C12N-015/00 ; C12Q-001/68
XPN - J63258580
XPR - 87JP-091976
it - CHICKEN DNA DISCRIMINATE HEN COCK LOCALISE FRAGMENT CONTAIN SPECIFIC NUCLEOTIDE SEQUENCE HEN SPECIFIC SEX CHROMOSOME
AB - J63258580 A DNA fragment containing a specific nucleotide sequence is localised in hen-specific sex chromosome (W chromosome).S USE/ADVANTAGE - Discrimination of male and female chickens is done by professionals. Recent progress of molecular biology can make the discrimination in molecular level possible. DNA probe is specific for female chick and thereby accurate discrimination of male and female chickens is possible. Because the DNA probe recognises hen or game cock specifically, judgement of unidentified young bird is possible. (11pp Dwg.No.0/0)
AW - DEOXYRIBONUCLEIC ACID
PAW - (NICH-) NICHIREI KK
TI - Chicken DNA - is used to discriminate between hen and cock, by localising fragment contg. specific nucleotide sequence in hen-specific sex chromosome
NC - 001
ORD - 1988-10-26

1/1 - (C) WPI / DERWENT
AN - 88-348692 [49]
AY - 88
XA - C88-154100
XP - N88-264156
DC - B04 C03 D16 P14
FS - CPI
PA - (NICH-) NICHIREI KK
PN - JP63258581 A 881026 DW8849
PR - JP870091977 870416
IC - A01K67/00 ; C07H21/04 ; C12N15/00 ; C12Q1/68
MC - B04-B04A1 B11-C08E B12-K04A6 C04-B04A1 C11-C08E C12-K04A6 D05-H01 D05-H09 D05-H12
M1 - [01] M423 M760 M903 N102 Q233 V600 V615 V754
- [02] B615 B701 B713 B720 B815 B831 C811 D011 D012 E720 J011 J171 J521 L921 M280 M314 M321 M332 M342 M372 M391 M423 M511 M520 M530 M540 M781 M903 N102 P831 Q233 Q444 Q505 V753; 00945
- [03] M423 M710 M750 M903 N102 Q233 V753
M6 - [04] M903 P831 Q233 Q505 R513 R515 R521 R536 R611 R614 R626 R627 R639
UP - 8849
UCC - 8849
AP - JP870091977 870416
XIC - A01K-067/00 ; C07H-021/04 ; C12N-015/00 ; C12Q-001/68
XPN - J63258581
XPR - 87JP-091977
it - DISCRIMINATE FEMALE MALE TURKEY DNA FRAGMENT CONTAIN SPECIFIC NUCLEOTIDE SEQUENCE LOCALISE TURKEY FEMALE SPECIFIC SEX CHROMOSOME
AB - J63258581 A DNA fragment contains a specific nucleotide sequence which is localised in turkey female-specific sex chromosome (W chromosome).
- (1) Turkey female-specific repeat sequence which exists in W chromosome emigrates at 0.4kb when electrophoresis is performed after digestion with an endonuclease, PstI. (2) Nucleotide sequence of PstI 0.4kb fragment comprises repeat sequence of about 20 base pairs in length. (3) The PstI 0.4kb fragment hybridises female turkey DNA but not with male turkey DNA.
- USE/ADVANTAGE - A DNA probe which is specific for female turkey and thereby accurate discrimination of male and female of turkey is possible. In addn., because the DNA probe recognises turkey specifically, judgement of unidentified young bird, whether it is turkey or not, is possible. (7pp Dwg.No.0/0)
AW - DEOXYRIBONUCLEIC ACID
PAW - (NICH-) NICHIREI KK
TI - Discriminating female and male turkey - using DNA fragment contg. specific nucleotide sequence localised in turkey female-specific sex chromosome
RRL - 00945
NC - 001
ORD - 1988-10-26

```
ID   GGJF11 .   standard; RNA; VRT; 2292 BP.
AC   D14316;
NI   g391639
DT   06-AUG-1993 (Rel. 36, Created)
DT   17-FEB-1994 (Rel. 38, Last updated, Version 2)
DE   delta-crystallin enhancer binding protein, complete sequence.
KW   ORF2.
OS   Gallus gallus (chicken)
OC   Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Aves;
OC   Neornithes; Neognathae; Galliformes; Phasianidae.
RN   [1]
RX   MEDLINE; 94116444.
RA   Funahashi J., Sekido R., Murai K., Kamachi Y., Kondoh H.;
RT   "Delta-crystallin enhancer binding protein deltaEF1 is a zinc
RT   finger-homeodomain protein implicated in postgastrulation

RT   embryogenesis";
RL   Development 119:433-446(1993).
RN   [2]
RP   1-2292
RA   Funahashi J.;
RT   ;
RL   Unpublished.
CC   Submitted (29-Jan-1993) to DDBJ by: Jun-ichi Funahashi Department
CC   of Molecular Biology School of Science Nagoya University Furo-cho,
CC   Chikusa-ku Nagoya 464-01 Japan Phone: 052-781-5111 x6684 Fax:
CC   052-782-8575
FH   Key             Location/Qualifiers
FH
FT   CDS               257. .1939
FT                     /note="beta-gal fusion protein binds to blocks 10 to 3 of
FT                     HN fragment of delta-crystallin enhancer."
FT                     /product="ORF2"
FT                     /codon_start=1

FT                     /db_xref="PID:g391640"
FT                     /translation="MELKKCCNHCYLIKPPDDNEFYNKQEALQHLIRSSGKLILLDKLL

FT                     IRLRERGNRVLIFSQMVRMLDILAEYLKYRQFPFQRLDGSIKGELRKQALDHFNAEGSE

FT                     DFCFLLSTRAGGLGINLASADTVVIFDSDWNPQNDLQAQARAHRIGQKKQVNIYRLVTK

FT                     GSVEEDILERAKKKMVLDHLVIQRMDTTGKTVLHTGSTPSSSTPFNKEELSAILKFGAE

FT                     ELFKEPEGEEQEPQEMDIDEILKRAETRENEPGPLTVGDELLSQFKVANFSNMDEDDIE

FT                     LEPERNSRNWEEIIPESQRRRIEEEERQKELEEIYMLPRMRNCAKQISFNGSEGRRSRS

FT                     RRYSGSDSDSITERKRPKKRGRPRTIPRENIKGFSDAEIRRFIKSYKKFGGPLERLDAV

FT                     ARDAELVDKSETDLRRLGELVHNGCIKALKDNSSGQERAGGRLGKVKGPTFRISGVQVN

FT                     AKLVISHEEELAPLHKSIPSDPEERKRYVIPCHTKAAHFDIDWGKEDDSNLLVGIYEYA

FT                     YGSWQCKVNGSRSQLNTEILPDDPDRTPRQNSYRPVQTTSLNY"
FT   source            1. .2292
FT                     /organism="Gallus gallus"
FT                     /dev_stage="13 day embryo"
FT                     /sequenced_mol="cDNA to mRNA"
FT                     /tissue_type="lens"
FT                     /clone_lib="lambda gt11" . . .


SCORES            Init1:  6327 Initn:  8063 Opt:  8924
            98.9% identity in 2293 bp overlap

              2200       2210       2220       2230       2240       2250
Sa1344 ATTTCATCATGCCAGAAAAATTTTCCTCCTGGGAAGATTTTGAAGAGGAGCATGGCAAAG
```


Ggjf11 CCTGGGGAAGATTTTGAAGAGGAGCATGGCAAAG
10 20 30
2260 2270 2280 2290 2300 2310
Sa1344 GAAGAGAGTATGGTTATGCAAGTCTTCACAAAGAGCTTGAACCATTTTTACTAAGAAGAG
Ggjf11 GAAGAGAGTATGGTTATGCAAGTCTTCACAAAGAGCTTGAACCATTTTTACTAAGAAGAG
40 50 60 70 80 90
2320 2330 2340 2350 2360 2370
Sa1344 TTAAAAAAGATGTAGAAAAGTCTTTACCTGCTAAGGTTGAACAAATTCTGAGGATGGAAA
Ggjf11 TTAAAAAAGATGTAGAAAAGTCTTTACCTGCTAA-GTTGAACAAATTCTGAGGATGGAAA
100 110 120 130 140 150
2380 2390 2400 2410 2420 2430
Sa1344 TGAGTGCATTGCAGAAGCAATATTACAAGTGGATTTTAACAAGGAATTATAAAGCCCTCA
Ggjf11 TGAGTGCATTGCAGAAGCAATATTACAAGTGGATTTTAACAAGGAATTATAAAGCCCTCA
160 170 180 190 200 210
2440 2450 2460 2470 2480 2490
Sa1344 GTAAAGGTTCAAAAGGCAGTACCTCAGGCTTTCTGAACATTATGATGGAACTTAAGAAGT
Ggjf11 GTAAAGGTTCAAAAGGCAGTACCTCAGGC-TTCTGAACATTATGATGGAACTTAAGAAGT
220 230 240 250 260 270
2500 2510 2520 2530 2540 2550
Sa1344 GTTGTAACCATTGCTACCTCATTAAGCCACCAGATGATAATGAATTCTATAATAAACAGG
Ggjf11 GTTGTAACCATTGCTACCTCATTAAGCCACCAGATGATAATGAATTCTATAATAAACAGG
280 290 300 310 320 330
2560 2570 2580 2590 2600 2610
Sa1344 AGGCCTTACAGCATTTGATACGTAGCAGCGGGAAACTAATCCTTCTTGACAAGCTACTGA
Ggjf11 AGGCCTTACAGCATTTGATACGTAGCAGCGGGAAACTAATCCTTCTTGACAAGCTACTGA
340 350 360 370 380 390
2620 2630 2640 2650 2660 2670
Sa1344 TTCGTCTGCGAGAACGTGGCAACAGAGTTCTGATTTTCTCTCAGATGGTGAGGATGCTGG
Ggjf11 TTCGTCTGCGAGAACGTGGCAACAGAGTTCTGATTTTCTCTCAGATGGTGAGGATGCTGG
400 410 420 430 440 450
2680 2690 2700 2710 2720 2730
Sa1344 ACATCCTAGCAGAATATCTGAAGTATCGCCAGTTTCCCTTCCAGAGACTTGATGGATCAA
Ggjf11 ACATCTTAGCAGAATATCTGAAGTATCGCCAGTTTCCCTTCCAGAGACTTGATGGATCAA
460 470 480 490 500 510
2740 2750 2760 2770 2780 2790
Sa1344 TAAAAGGGGAATTGAGGAAGCAAGCACTGGATCATTTCAATGCAGAAGGATCAGAGGATT
Ggjf11 TAAAAGGGGAATTGAGGAAGCAAGCACTGGATCATTTCAATGCAGAAGGATCAGAGGATT
520 530 540 550 560 570
2800 2810 2820 2830 2840 2850
Sa1344 TCTGTTTTTTACTGTCTACAAGAGCTGGAGGATTAGGTATTAACTTGGCATCTGCTGACA
Ggjf11 TCTGTTTTTTACTGTCTACAAGAGCTGGAGGATTAGGTATTAACTTGGCATCTGCTGACA

```
          580        590        600        610        620        630

          2860       2870       2880       2890       2900       2910
Sa1344 CTGTAGTTATTTTTGATTCTGACTGGAATCCACAGAATGATCTGCAGGCACAGGCGAGAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 CTGTAGTTATTTTTGATTCTGACTGGAATCCACAGAATGATCTGCAGGCACAGGCGAGAG
          640        650        660        670        680        690

          2920       2930       2940       2950       2960       2970
Sa1344 CTCATAGAATTGGACAGAAGAAACAGGTTAATATTTATCGGCTAGTCACAAAAGGATCAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 CTCATAGAATTGGACAGAAGAAACAGGTTAATATTTATCGGCTAGTCACAAAAGGATCAG
          700        710        720        730        740        750

          2980       2990       3000       3010       3020       3030
Sa1344 TAGAAGAAGATATTCTTGAAAGAGCCAAGAAGAAGATGGTGCTAGACCATTTAGTAATTC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 TAGAAGAAGATATTCTTGAAAGAGCCAAGAAGAAGATGGTGCTAGACCATTTAGTAATTC
          760        770        780        790        800        810

          3040       3050       3060       3070       3080       3090
Sa1344 AGAGAATGGACACGACAGGAAAAACTGTTCTGCATACAGGTTCAACTCCATCAAGCTCTA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AGAGAATGGACACGACAGGAAAAACTGTTCTGCATACAGGTTCAACTCCATCAAGCTCTA
          820        830        840        850        860        870

          3100       3110       3120       3130       3140       3150
Sa1344 CACCTTTTAATAAAGAAGAGTTATCAGCTATTTTGAAGTTTGGTGCTGAGGAACTCTTTA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 CACCTTTTAATAAAGAAGAGTTATCAGCTATTTTGAAGTTTGGTGCTGAGGAACTCTTTA
          880        890        900        910        920        930

          3160       3170       3180       3190       3200       3210
Sa1344 AAGAACCTGAAGGAGAAGAACAGGAGCCCCAGGAAATGGATATAGATGAAATCTTGAAGA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AAGAACCTGAAGGAGAAGAACAGGAGCCCCAGGAAATGGATATAGATGAAATCTTGAAGA
          940        950        960        970        980        990

          3220       3230       3240       3250       3260       3270
Sa1344 GAGCTGAAACTCGGGAAAATGAGCCAGGTCCATTGACTGTAGGGGATGAGTTGCTTTCAC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 GAGCTGAAACTCGGGAAAATGAGCCAGGTCCATTGACTGTAGGGGATGAGTTGCTTTCAC
         1000       1010       1020       1030       1040       1050

          3280       3290       3300       3310       3320       3330
Sa1344 AGTTCAAGGTGGCGAACTTTTCCAATATGGATGAAGATGATATTGAGTTGGAACCAGAAA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AGTTCAAGGTGGCGAACTTTTCCAATATGGATGAAGATGATATTGAGTTGGAACCAGAAA
         1060       1070       1080       1090       1100       1110

          3340       3350       3360       3370       3380       3390
Sa1344 GAAATTCAAGAAATTGGGAAGAAATCATCCCAGAATCCCAACGGAGAAGGATAGAGGAGG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 GAAATTCAAGAAATTGGGAAGAAATCATCCCAGAATCCCAACGGAGAAGGATAGAGGAGG
         1120       1130       1140       1150       1160       1170

          3400       3410       3420       3430       3440       3450
Sa1344 AGGAAAGACAAAAAGAACTTGAAGAAATATACATGCTCCCGAGGATGAGAAACTGTGCAA
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AGGAAAGACAAAAAGAACTTGAAGAAATATACATGCTCCCGAGGATGAGAAACTGTGCAA
         1180       1190       1200       1210       1220       1230

          3460       3470       3480       3490       3500       3510
```

```
Sa1344 AACAGATCAGCTTTAATGGGAGTGAAGGAAGACGCAGTAGGAGCAGAAGATATTCTGGAT
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AACAGATCAGCTTTAATGGGAGTGAAGGAAGACGCAGTAGGAGCAGAAGATATTCTGGAT
           1240       1250       1260       1270       1280       1290

            3520       3530       3540       3550       3560       3570
Sa1344 CTGATAGTGACTCCATCACAGAAAGAAAACGGCCAAAAAAGCGTGGAAGACCTCGAACCA

       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 CTGATAGTGACTCCATCACAGAAAGAAAACGGCCAAAAAAGCGTGGAAGACCTCGAACCA
           1300       1310       1320       1330       1340       1350

            3580       3590       3600       3610       3620       3630
Sa1344 TTCCTCGAGAAAATATTAAAGGATTTAGTGATGCAGAGATCAGGCGGTTTATCAAGAGTT
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 TTCCTCGAGAAAATATTAAAGGATTTAGTGATGCAGAGATCAGGCGGTTTATCAAGAGTT
           1360       1370       1380       1390       1400       1410

            3640       3650       3660       3670       3680       3690
Sa1344 ACAAGAAATTTGGTGGCCCTCTGGAAAGGTTAGATGCTGTAGCTAGAGATGCTGAACTGG

       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 ACAAGAAATTTGGTGGCCCTCTGGAAAGGTTAGATGCTGTAGCTAGAGATGCTGAACTGG
           1420       1430       1440       1450       1460       1470

            3700       3710       3720       3730       3740       3750
Sa1344 TTGATAAATCTGAGACAGACCTTAGACGTTTGGGTGAACTTGTACATAATGGATGCATTA
       |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Ggjf11 TTGATAAATCTGAGACAGACCTTAGACGTTTGGGTGAGCTTGTACATAATGGATGCATTA
           1480       1490       1500       1510       1520       1530

            3760       3770       3780       3790       3800       3810
Sa1344 AGGCTTTAAAGGACAATTCATCTGGACAAGAAAGAGCAGGAGGTAGACTTGGGAAAGTTA

       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AGGCTTTAAAGGACAATTCATCTGGACAAGAAAGAGCAGGAGGTAGACTTGGGAAAGTTA
           1540       1550       1560       1570       1580       1590

            3820       3830       3840       3850       3860       3870
Sa1344 AAGGCCCAACGTTTCGAATCTCAGGAGTGCAGGTGAATGCAAAACTAGTCATCTCTCACG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AAGGCCCAACGTTTCGAATCTCAGGAGTGCAGGTGAATGCAAAACTAGTCATCTCTCACG
           1600       1610       1620       1630       1640       1650

            3880       3890       3900       3910       3920       3930
Sa1344 AAGAAGAGCTGGCACCACTGCACAAATCCATTCCTTCAGATCCAGAAGAAAGGAAAAGAT

       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AAGAAGAGCTGGCACCACTGCACAAATCCATTCCTTCAGATCCAGAAGAAAGGAAAAGAT
           1660       1670       1680       1690       1700       1710

            3940       3950       3960       3970       3980       3990
Sa1344 ATGTCATCCCATGCCACACCAAGGCTGCTCACTTCGATATAGATTGGGGTAAAGAAGATG
       | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 ACGTCATCCCATGCCACACCAAGGCTGCTCACTTCGATATAGATTGGGGTAAAGAAGATG
           1720       1730       1740       1750       1760       1770

            4000       4010       4020       4030       4040
Sa1344 ATTCCAATCTGTTAGTAGGCATCTATGAATATGGCTATGGCAGCTGGGAAATGATAA--A

       |||||||||||||||||||||||||||||||| |||||||||| || |||| || |
Ggjf11 ATTCCAATCTGTTAGTAGGCATCTATGAATATGCCTATGGCAGCT-GGCAATGTAAAGTA
           1780       1790       1800       1810       1820       1830

      4050       4060       4070       4080       4090       4100
Sa1344 AATGGATCCAGATCTCAGCTTAACACAGAAGATTTTACCTGATGATCCAGACAAGAAACC
       |||||||||||||||||||||||||||||| |||||||||||||||||||| |||| |
Ggjf11 AATGGATCCAGATCTCAGCTTAACACAG-AGATTTTACCTGATGATCCAGAC-AGAACTC
```

```
      1840       1850       1860       1870       1880

  4110       4120       4130       4140       4150       4160
Sa1344 CCAGGCAAAGCAGCTACAGACCCGTGCAGACTACCTCATTAAATTACTGAATAAAGACCT

       | ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 CAAGGCAAAACAGCTACAGACCCGTGCAGACTACCTCATTAAATTACTGAATAAAGACCT
    1890       1900       1910       1920       1930       1940

     4170       4180       4190       4200       4210       4220
Sa1344 TGCAAGAAAGGAAG-CAC-AAAGGCTTGCTGGTGCAGGCAATTCCAAGAGAAGGAAGACA
       |||||||||||||||| | | ||||||||||||||||||||||||||||||||||||||
Ggjf11 TGCAAGAAAGGAAGTCGCTAAAGGCTTGCTGGTGCAGGCAATTCCAAGAGAAGGAAGACA
    1950       1960       1970       1980       1990       2000

       4230       4240       4250       4260       4270       4280
Sa1344 AGAAATAAGAAGAATAAGATGAAGGCTTCAAAAATAAAAGAAGAAATAAAGAGTGATTCT

       || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 AGGAATAAGAAGAATAAGATGAAGGCTTCAAAAATAAAAGAAGAAATAAAGAGTGATTCT
    2010       2020       2030       2040       2050       2060

       4290       4300       4310       4320       4330       4340
Sa1344 TCACCACAACCCTCAGAAAAATCTGATGAAGATGATGAGGAGGAGGATAACAAGGTAAAT
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 TCACCACAACCCTCAGAAAAATCTGATGAAGATGATGAGGAGGAGGATAACAAGGTAAAT
    2070       2080       2090       2100       2110       2120

       4350       4360       4370       4380       4390       4400
Sa1344 GAAATGAAATCTGAAAATAAAGAAAAATCTAAAAAAATTCCATTGCTGGATACTCCAGTT

       ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Ggjf11 GAAATCAAATCTGAAAATAAAGAAAAATCTAAAAAAATTCCATTGCTGGATACTCCAGTT
    2130       2140       2150       2160       2170       2180

       4410       4420       4430       4440       4450       4460
Sa1344 CATATTACTGCAACCAGTGAACCAGTTCCTATCTCAGAAGAATCTGAAGAACTCCATCAG
       |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Ggjf11 CATATTACTGCAACCAGTGAACCAGTTCCTATCTCAGAAGAATCTGAAGAACTGGATCAG
    2190       2200       2210       2220       2230       2240

       4470       4480       4490       4500       4510       4520
Sa1344 AAGACATTTAGTGTGTGCAAAGAAAGAATGAGGCCTGTCAAAGCAGCACTGAAACAGCTG

       |||||||||||||||||||||||||||||||||||||||
Ggjf11 AAGACATTTAGTGTGTGCAAAGAAAGAATGAGGCCTGTCAAAG
    2250       2260       2270       2280       2290

       4530       4540       4550       4560       4570       4580
Sa1344 GATAGACCAGAGAAGGGCCTTTCTGAAAGGGAGCAGCTGGAACATACTAGGCAGTGTCTA
```